

OPTIMAL PERTURBATIONS OF RUNGE-KUTTA METHODS

INMACULADA HIGUERAS*, DAVID I. KETCHESON†, AND TIHAMÉR A. KOCSIS‡

Abstract. Perturbed Runge–Kutta methods (also referred to as downwind Runge–Kutta methods) can guarantee monotonicity preservation under larger step sizes relative to their traditional Runge–Kutta counterparts. In this paper we study, the question of how to optimally perturb a given method in order to increase the radius of absolute monotonicity (a.m.). We prove that for methods with zero radius of a.m., it is always possible to give a perturbation with positive radius. We first study methods for linear problems and then methods for nonlinear problems. In each case, we prove upper bounds on the radius of a.m., and provide algorithms to compute optimal perturbations. We also provide optimal perturbations for many known methods.

Key words. Strong Stability Preserving, Monotonicity, Runge-Kutta methods

AMS subject classifications. 65L06, 65L20, 65M20

1. Introduction. Strong stability preserving Runge–Kutta (RK) methods were first introduced by Shu and Osher [23] in the context of time integration for first-order hyperbolic conservation laws:

$$\mathcal{U}_t + \mathcal{F}(\mathcal{U})_x = 0, \quad \mathcal{U}(x, t = 0) = \mathcal{U}_0. \quad (1.1)$$

After semi-discretization, (1.1) takes the form of an initial-value ordinary differential equation system:

$$u'(t) = f(u) \quad u(0) = u_0, \quad (1.2)$$

where f is a discrete approximation to $-\mathcal{F}$. In the scalar case \mathcal{U} is dissipative, and it is natural to seek a semi-discretization that is dissipative:

$$\frac{d}{dt} \|u\| \leq 0, \quad (1.3)$$

where $\|\cdot\|$ denotes a convex functional (e.g., a norm, a semi-norm, \dots). This is achieved by biasing the discretization f in the upwind direction. A necessary condition for (1.3) is monotonicity under an explicit Euler step [18, p. 501]:

$$\|v + hf(v)\| \leq \|v\|, \quad \text{for all } v, \text{ and for } h \text{ satisfying } 0 \leq h \leq h_0, \quad (1.4)$$

where $h_0 > 0$ (in general h_0 may depend on v). Let u_n, u_{n+1} denote approximations, computed by some numerical integrator, to the solution at successive time steps t_n and $t_{n+1} = t_n + h$. Under

*Public University of Navarre, Pamplona 31006, Spain (higueras@unavarra.es). Supported by Ministerio de Economía y Competitividad, Spain, Project MTM2011-23203

†King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia (david.ketcheson@kaust.edu.sa). Supported by KAUST Award No. FIC/2010/052000000231

‡Széchenyi István University, Győr, H-9026, Hungary (katihi@sze.hu). Supported by KAUST Award No. FIC/2010/052000000231 and by TÁMOP-4.2.2.A-11/1/KONV-2012-0012: Basic research for the development of hybrid and electric vehicles. The Project is supported by the Hungarian Government and co-financed by the European Social Fund.

the forward Euler monotonicity condition (1.4), it is possible to prove that many Runge–Kutta and linear multistep methods also give monotone solutions; i.e., solutions that satisfy

$$\|u_{n+1}\| \leq \|u_n\| \quad \text{for } h \text{ satisfying } 0 \leq h \leq Rh_0. \quad (1.5)$$

Such methods are known as strong stability preserving (SSP) methods, and the factor R is known as the radius of absolute monotonicity or SSP coefficient of the method. SSP methods necessarily have non-negative coefficients, since the monotonicity property is proved using (1.4) and convexity.

Monotonicity cannot be ensured using only assumption (1.4) in general for methods with negative coefficients [18, Thm. 4.2], or even for some methods (such as the classical fourth-order RK method) with non-negative coefficients [18, Thm. 9.6]. In order to accommodate such methods, a second discrete approximation to $-\mathcal{F}$ is introduced and referred to as \tilde{f} . This discretization must be monotone under an explicit Euler step with negative step size:

$$\|v - hf(v)\| \leq \|v\|, \quad \text{for all } v, \text{ and for } h \text{ satisfying } 0 \leq h \leq \tilde{h}_0, \quad (1.6)$$

where $\tilde{h}_0 > 0$. In the context of hyperbolic problems, \tilde{f} must be biased in the downwind direction, and typically $\tilde{h}_0 = h_0$. The downwind spatial discretization \tilde{f} is to be used in place of f wherever a negative coefficient appears in the time integration method, in order to ensure monotonicity of the overall method. Introduction of a downwind discretization makes it possible to ensure monotonicity for a broader class of methods, including the classical RK method of order four. It also makes it possible to ensure monotonicity for many methods under larger step sizes.

Methods that use both upwind and downwind operators can naturally be viewed as *perturbed Runge–Kutta methods*. Although they are also connected to additive RK methods (see [9, 10]), in the present work we will employ the perturbation viewpoint, and refer to methods that use downwind discretization as perturbed RK methods.

During the last quarter century, a number of additional authors have studied monotonicity for methods that use downwind discretization. The main motivation for this work has been to break the “order barrier” that restricts explicit RK methods to order four, or to find new methods with larger SSP coefficients. In this context, numerical optimization of the SSP coefficient for RK methods with negative coefficients was conducted for explicit methods in [22, 21, 6] and for implicit methods in [16]. In each case, optimization was carried out over methods with a specified order and number of stages.

The present work stems from a different motivation. Monotonicity preservation is not the only numerical property of interest in applications, and practitioners may wish to use a particular integrator that has small or zero SSP coefficient. Our goal is then to perturb the prescribed method in order to achieve larger monotonicity-preserving timesteps. Little work has been done in this direction, because it is not known how to find the best perturbation for a given method. That problem is the main focus of this work. Most of our results concern only explicit methods, although one major result (Theorem 3.1) pertains also to implicit methods.

1.1. Perturbed Runge–Kutta methods. A Runge–Kutta method applied to the initial value problem (1.2) computes approximations $u_n \approx u(t_n)$ by

$$Y = u_n e + hKF, \quad (1.7a)$$

$$u_{n+1} = Y_{s+1}. \quad (1.7b)$$

Here s is the number of stages, e is a vector whose entries are equal to one, Y is the vector containing the stage values and the numerical solution, $Y = (Y_1, \dots, Y_s, Y_{s+1})^t$, $[F]_i = f(Y_i)$, and K is the $(s+1) \times (s+1)$ matrix of Butcher coefficients:

$$K = \begin{pmatrix} A & 0 \\ b^t & 0 \end{pmatrix}.$$

In this work we study perturbed Runge–Kutta methods:

$$Y = u_n e + hKF + h\tilde{K}(F - \tilde{F}), \quad (1.8a)$$

$$u_{n+1} = Y_{s+1}, \quad (1.8b)$$

where \tilde{K} is given by

$$\tilde{K} = \begin{pmatrix} \tilde{A} & 0 \\ \tilde{b}^t & 0 \end{pmatrix},$$

and \tilde{F} is defined analogously to F . We assume that the perturbation \tilde{A} has the same structure (strictly lower-triangular, lower-triangular, or full) as the matrix A .

Observe that the perturbed method (1.8) reduces to the RK method (1.7) when $\tilde{F} = F$. Method (1.8) may be viewed as approximating the solution of the perturbed problem

$$u'(t) = f(u) + (f(u) - \tilde{f}(u)),$$

where $\tilde{f} \approx f$.

We assume that f and \tilde{f} satisfy the explicit Euler assumptions (1.4) and (1.6), respectively, with $\tilde{h}_0 = h_0$.

Most previous works, including [22, 21], have focused on methods with the following property

Property C:

We say that a perturbation \tilde{K} to an RK method K possesses *property C* if, for each value of j

$$\tilde{K}_{ij} \neq 0 \text{ (for some } i) \implies K_{ij} = 0 \text{ (for all } i). \quad (1.9)$$

In words, property C means that in the j th column, only one of K, \tilde{K} has any nonzero entries. Thus, only one of $f(y_j), \tilde{f}(y_j)$ need ever be evaluated, so only s total function evaluations are required per step. In [6] it was shown that for WENO discretizations, the cost of computing both $f(y_j)$ and $\tilde{f}(y_j)$ is much less than twice the cost of computing $f(y_j)$ alone. Therefore methods that without property C may also be of practical interest. In the present work, we do not assume property C.

1.2. Scope and outline. The central question of the present work is

- Given a fixed RK method, what perturbation results in the largest radius of absolute monotonicity for the perturbed method?

We investigate this question in the context of both linear problems (Section 2) and nonlinear problems (Section 3).

For explicit methods applied to linear problems, the question above can be cast in terms of absolute monotonicity of the (bivariate) stability polynomial. In Section 2.1.1 we prove a general upper bound on the radius of absolute monotonicity of the stability polynomial of an explicit

perturbed RK method with s stages and linear order p . In Section 2.1.2, we provide an algorithm for computing tighter bounds, and tabulate some of the resulting numerical values. Examples of optimal methods are given in 2.2.

Section 3 is devoted to perturbed Runge–Kutta methods for nonlinear problems. Theorem 3.1 states that a perturbation with positive radius of absolute monotonicity exists for every Runge–Kutta method. Theorems 3.3 and 3.4 give simple upper bounds on the optimal perturbed radius for a given explicit method. In Section 3.5 we give two algorithms for computing optimal perturbations. The first is provably correct but approximate, while the second is heuristic but exact and agrees with the first in all cases we have tested. Both are applicable only to explicit methods. These algorithms have been implemented in the software package Nodepy [17]. We conclude Section 3 with an application of the theorems and algorithms to optimal perturbations of some RK methods from the literature. Among the results is the first truly optimal perturbation for the classical 4th-order method of Kutta.

Section 4 contains some conclusions as well as some open questions to be studied in the future.

Finally, in the Appendix we give some details on perturbations for the family of second order 2-stage methods and the classical fourth order RK method.

2. Explicit perturbed Runge–Kutta methods for linear problems. To study the behavior of the perturbed Runge–Kutta method (1.8) for linear problems, we apply it to a linear scalar test problem, setting $f(u) = \lambda u$ and $\tilde{f}(u) = \tilde{\lambda} u$ in (1.8). This results in the iteration

$$u_{n+1} = \phi_{(K, \tilde{K})}(z, -\tilde{z}) u_n,$$

where $z = h\lambda$, $\tilde{z} = h\tilde{\lambda}$ and

$$\phi_{(K, \tilde{K})}(z, \tilde{z}) = 1 + \left(zb^t + (z + \tilde{z})\tilde{b}^t \right) \left(I - zA - (z + \tilde{z})\tilde{A} \right)^{-1} e. \quad (2.1)$$

We refer to (2.1) as the *stability function of the perturbed Runge–Kutta method* (1.8).

We say that a function $\psi : \mathbb{R} \rightarrow \mathbb{R}$ is absolutely monotonic (a.m.) at ξ if all derivatives at ξ exist and they are non-negative. For a function $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$ the concept of absolute monotonicity can be defined in a similar way: ψ is a.m. at $(\xi, \tilde{\xi})$ if all derivatives at $(\xi, \tilde{\xi})$ exist and they are non-negative.

Given a function $\psi(z, \tilde{z})$, we define the *radius of absolute monotonicity* as

$$R(\psi) = \sup \{ r \in \mathbb{R} \mid r = 0, \text{ or } r > 0, \text{ and } \psi(z, \tilde{z}) \text{ is a.m. at } (-r, -r) \}. \quad (2.2)$$

For a perturbed Runge–Kutta method (1.8) with coefficients (K, \tilde{K}) , we write $R_{\text{Lin}}(K, \tilde{K})$ to denote the radius of absolute monotonicity of its stability function:

$$R_{\text{Lin}}(K, \tilde{K}) = R(\phi_{(K, \tilde{K})}).$$

The quantity $R_{\text{Lin}}(K, \tilde{K})$ is referred to as the *threshold factor* due to its role in the step size for monotonicity. For a given Runge–Kutta method (1.7) with coefficients K , we are interested in determining perturbations \tilde{K} that give the largest threshold factor. The corresponding *threshold factor of the optimal perturbation* is denoted by

$$R_{\text{Lin}}^{\text{opt}}(K) = \sup_{\tilde{K}} R_{\text{Lin}}(K, \tilde{K}). \quad (2.3)$$

The supremum in (2.3) is taken over all strictly lower triangular matrices \tilde{K} , in order to preserve the explicit nature of the method. A perturbation \tilde{K} such that

$$R_{\text{Lin}}(K, \tilde{K}) = R_{\text{Lin}}^{\text{opt}}(K)$$

will be called an *optimal perturbation of the method K for the linear problem*.

Taking $\tilde{K} = 0$ gives a (not perturbed) Runge–Kutta method (1.7) and a (not perturbed) stability function ϕ_K . In this case we denote the threshold factor $R_{\text{Lin}}(K, 0)$ simply by $R(\phi_K)$. Clearly

$$R(\phi_K) \leq R_{\text{Lin}}^{\text{opt}}(K). \quad (2.4)$$

In the next section, we give upper bounds on $R_{\text{Lin}}^{\text{opt}}(K)$.

2.1. Upper bounds on the threshold factor for optimal perturbations . For any explicit perturbed RK method (K, \tilde{K}) of linear order p , it can be seen that $\phi_{(K, \tilde{K})} \in \tilde{\Pi}_{s,p}$, where $\tilde{\Pi}_{s,p}$, with $p \leq s$, denotes the set of bivariate polynomials with the following properties:

1. $\psi(z, \tilde{z}) = \sum_{j=0}^p \frac{z^j}{j!} + \sum_{j=p+1}^s \sigma_j z^j + (z + \tilde{z})\Psi(z, \tilde{z})$;
2. Ψ is a polynomial of combined degree at most $s - 1$.

In this section we investigate

$$\tilde{R}_{s,p} = \sup \left\{ R(\psi) \mid \psi(z, \tilde{z}) \in \tilde{\Pi}_{s,p} \right\},$$

Clearly

$$R_{\text{Lin}}^{\text{opt}}(K) \leq \tilde{R}_{s,p}. \quad (2.5)$$

However, not all functions in $\tilde{\Pi}_{s,p}$ can be realized as the stability function of an s -stage perturbed Runge–Kutta method (1.8), so inequality (2.5) is often strict (see Example 2.3 below). In case the optimal polynomial is realizable, the corresponding method may be of interest for the integration of linear systems.

In Subsection 2.1.1 we give an upper bound for $\tilde{R}_{s,p}$. In Subsection 2.1.2, we give an algorithm to compute, $\tilde{R}_{s,p}$ for given s and p , along with numerical values.

2.1.1. Upper bound on $\tilde{R}_{s,p}$.

LEMMA 2.1. *Let $\varphi(z)$ be a polynomial satisfying*

$$\begin{aligned} \varphi(z) &= 1 + \gamma_1 z + \cdots + \gamma_p z^p + \gamma_{p+1} z^{p+1} + \cdots + \gamma_s z^s \\ \gamma_j &\geq \frac{1}{j!}, \quad j = 1, \dots, p. \end{aligned} \quad (2.6)$$

Then the radius of absolute monotonicity of φ satisfies

$$R(\varphi) \leq \sqrt[p]{s(s-1) \cdots (s-p+1)}. \quad (2.7)$$

Proof. If $R(\varphi) = 0$, inequality (2.7) is trivial. Let $\varphi(z)$ satisfy (2.6) and be absolutely monotonic at $-r$ with $r > 0$. Then it can be written as

$$\varphi(z) = \sum_{j=0}^s \alpha_j \left(1 + \frac{z}{r}\right)^j = \sum_{j=0}^s \alpha_j \left(\sum_{\ell=0}^j \frac{z^\ell}{r^\ell} \binom{j}{\ell} \right) = \sum_{\ell=0}^s \left(\sum_{j=\ell}^s \alpha_j \binom{j}{\ell} \right) \frac{z^\ell}{r^\ell},$$

where $\alpha_j \geq 0$, and $\sum_j \alpha_j = 1$. As φ is of the form (2.6), the coefficient of z^p is larger than $1/p!$. Some computations give

$$\frac{1}{p!} \leq \left(\sum_{j=p}^s \alpha_j \binom{j}{p} \right) \frac{1}{r^p} \leq \left(\sum_{j=p}^s \alpha_j \right) \binom{s}{p} \frac{1}{r^p} \leq \binom{s}{p} \frac{1}{r^p} = \frac{s(s-1) \cdots (s-p+1)}{p! r^p}.$$

Consequently,

$$r \leq \sqrt[p]{s(s-1) \cdots (s-p+1)}.$$

□

We remark that equality in (2.7) is obtained for the polynomial

$$\varphi(z) = \left(1 + \frac{z}{r}\right)^s, \quad (2.8)$$

where $r = \sqrt[p]{s(s-1) \cdots (s-p+1)}$.

With the previous results, we can prove an upper bound on $\tilde{R}_{s,p}$.

THEOREM 2.2.

$$\tilde{R}_{s,p} \leq \sqrt[p]{s(s-1) \cdots (s-p+1)}. \quad (2.9)$$

Proof. If $R(\psi) = 0$ for all $\psi \in \tilde{\Pi}_{s,p}$, then $\tilde{R}_{s,p} = 0$ and inequality (2.9) is true. Otherwise, there exists a function $\psi \in \tilde{\Pi}_{s,p}$ a.m. at $(-r, -r)$ with $r > 0$. By [10, Lemmas 2.9 and 2.10], ψ is a.m. at the points (ξ, ξ) , with $\xi \in [-r, 0]$. Writing $\psi(z, \bar{z}) = \sum \sum \mu_{jk} z^j \bar{z}^k$ and differentiating shows that all coefficients μ_{jk} are non-negative since ψ is a.m. at $(0, 0)$. Thus $\psi(z, z)$ (viewed as a function of one variable) is of the form (2.6) and is a.m. at $-r$. Application of Lemma 2.1 gives the desired result. □

2.1.2. Numerical computation of bounds $\tilde{R}_{s,p}$. In this section we provide a means to compute tighter bounds on $\tilde{R}_{s,p}$ for given s, p , using linear programming. The material in this section closely follows [15, Section 4.6.2]. The stability function (2.1) of an explicit s -stage perturbed Runge–Kutta method (1.8) with linear order p can also be written in the form

$$\psi(z, \bar{z}) = \sum_{j=0}^s \sum_{\ell=0}^j \gamma_{j\ell} \left(1 + \frac{z}{r}\right)^{j-\ell} \left(1 + \frac{\bar{z}}{r}\right)^\ell \quad \text{with } \gamma_{j\ell} = \frac{r^j}{j!} \frac{\partial^j \psi_i}{\partial z^{j-\ell} \partial \bar{z}^\ell}; \quad (2.10)$$

furthermore, $\psi(z, -z) = \phi_K(z) = \exp(z) + \mathcal{O}(z^{p+1})$. After considerable manipulation we find that $\psi(z, -z) = \sum_{i=0}^s C_i z^i$ where

$$C_i(r, \gamma) = \sum_{j=i}^s \sum_{\ell=0}^j \gamma_{j\ell} \sum_{m=\max(0, i-\ell)}^{\min(i, j-\ell)} \binom{j-\ell}{m} \binom{\ell}{i-m} \frac{(-1)^{i-m}}{r^i}$$

TABLE 2.1
 $\tilde{R}_{s,p}$: upper bounds on the threshold factors for optimal perturbations

s p	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	2.00	1.41								
3	3.00	2.45	1.60							
4	4.00	3.46	2.49	2.00						
5	5.00	4.47	3.20	2.94	2.18					
6	6.00	5.48	4.00	3.65	3.11	2.58				
7	7.00	6.48	4.86	4.45	3.88	3.55	2.76			
8	8.00	7.48	5.77	5.31	4.57	4.32	3.72	3.15		
9	9.00	8.49	6.62	6.22	5.24	5.02	4.52	4.14	3.33	
10	10.00	9.49	7.42	7.09	5.95	5.70	5.25	4.96	4.32	3.73

Hence we have the following problem for existence of a polynomial (2.10) with perturbed threshold factor at least r and order at least p :

Given r find γ such that

$$\gamma_{j\ell} \geq 0 \quad 0 \leq \ell \leq j \leq s \quad (2.11a)$$

$$C_i(r, \gamma) = \frac{1}{i!} \quad 0 \leq i \leq p. \quad (2.11b)$$

Since (2.11b) is a system of linear equations (in γ) then for any given value of r (2.11) represents a linear programming feasibility problem. Hence we can use bisection and an LP solver to find $\tilde{R}_{s,p}$, as was done for similar problems in [13, 14]. Table 2.1 gives computed values of $\tilde{R}_{s,p}$ for s and p up to ten.

2.2. Examples. We now give some examples of optimal polynomials and Runge–Kutta methods.

2.2.1. Polynomials achieving $\tilde{R}_{s,p}$. The algorithm just described also provides coefficients for an optimal polynomial, which may or may not be realizable as the stability function of a perturbed Runge–Kutta method.

The optimal first-order polynomial for any s is just the stability polynomial of a (not perturbed) Runge–Kutta method consisting of s repeated forward Euler steps.

The optimal order-two polynomial of degree s also has a simple form:

$$\psi_{s,2}(z, \tilde{z}) = \frac{2(s+r)-1}{2(s+r)} \left(1 + \frac{z}{r}\right)^s + \frac{1}{2(s+r)} \left(1 + \frac{\tilde{z}}{r}\right)^s, \quad (2.12)$$

where $r = \tilde{R}_{s,2} = \sqrt{s(s-1)}$. Observe that bound (2.9) is sharp for $p = 2$.

Some of the other optimal polynomials also have rational coefficients. Two optimal degree-four fourth order polynomials we found are

$$\psi_{4,4}^1(z, \tilde{z}) = \frac{1}{3} \left(1 + \frac{z}{r}\right)^2 + \frac{17}{48} \left(1 + \frac{z}{r}\right)^4 + \frac{14}{48} \left(1 + \frac{z}{r}\right)^2 \left(1 + \frac{\tilde{z}}{r}\right)^2 + \frac{1}{48} \left(1 + \frac{\tilde{z}}{r}\right)^4,$$

and

$$\psi_{4,4}^2(z, \tilde{z}) = \frac{7}{16} \left(1 + \frac{z}{r}\right)^4 + \frac{3}{8} \left(1 + \frac{z}{r}\right)^2 \left(1 + \frac{\tilde{z}}{r}\right)^2 + \frac{1}{6} \left(1 + \frac{z}{r}\right)^3 \left(1 + \frac{\tilde{z}}{r}\right) + \frac{1}{48} \left(1 + \frac{\tilde{z}}{r}\right)^4,$$

where $r = \tilde{R}_{4,4} = 2$. Thus the optimal polynomial in $\tilde{\Pi}_{s,p}$ is in general not unique.

We have also computed (not shown in this paper) the exact values of $\tilde{R}_{s,p}$ and polynomials $\psi_{s,p}$ for $p = 3$ and $s = 3, 4, 5, 6$. In each case, we found an optimal polynomial of the form (2.10) where p of the coefficients γ_{sj} are non-zero and the coefficients γ_{ij} with $i < s$ are all zero.

Exact values of $\tilde{R}_{s,p}$ can be found in a systematic way as follows. First, a high-precision approximation can be computed using bisection and linear programming as described above. In practice, this yields a set of coefficients γ_{jl} in which only p values are non-zero [7]. Setting the remaining values to zero *a priori* in (2.11b) yields a system of $p+1$ equations in p unknowns which, nevertheless, possesses a solution. The solution may be found using a symbolic linear algebra package. We do not pursue this further in the present work.

REMARK 1. *As noted already, not all polynomials of the form (2.10) can be realized as the stability function of a perturbed Runge–Kutta method (1.8) with s stages. For example, the polynomial (2.12) with $s = 2$ is not the stability function of any two-stage method (i.e. using only evaluations of $f(u_n), \tilde{f}(u_n), f(y_1), \tilde{f}(y_1)$). It can be realized as the stability function of a method that has three stages, using evaluations of $f(u_n), \tilde{f}(u_n), f(y_1), \tilde{f}(y_2)$. The difference in cost between such methods depends on the nature of f, \tilde{f} ; see [6]. For this reason, we stress that the values in Table 2.1 are only upper bounds on what can be achieved. We do not pursue the topic further here.* \square

2.2.2. Optimal threshold factors for perturbations of specified 2-stage 2nd-order methods. We have no general method for finding $R_{\text{Lin}}^{\text{opt}}(K)$ nor a corresponding method. In this section we report results of some symbolic searches. In the case of the second-order methods, due to the small number of free parameters, it is not difficult to prove that the results below are truly optimal.

EXAMPLE 2.3. *We consider explicit perturbed second-order 2-stage methods*

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \alpha & \alpha & 0 \\ \hline K & 1 - \frac{1}{2\alpha} & \frac{1}{2\alpha} \end{array} \quad \begin{array}{c|cc} 0 & 0 \\ \tilde{a}_{21} & 0 \\ \hline \tilde{K} & \tilde{b}_1 & \tilde{b}_2 \end{array}. \quad (2.13)$$

For these methods, function (2.1) can be expanded as

$$\phi_{(K,\tilde{K})}(z, \tilde{z}) = 1 + z + \frac{1}{2}z^2 + \beta_{11}z(z + \tilde{z}) + \beta_1(z + \tilde{z}) + \beta_2(z + \tilde{z})^2, \quad (2.14)$$

where

$$\beta_{11} = b^t \tilde{A}e + \tilde{b}^t Ae = \tilde{b}_2 a_{21} + b_2 \tilde{a}_{21}, \quad \beta_1 = \tilde{b}^t e = \tilde{b}_1 + \tilde{b}_2, \quad \beta_2 = \tilde{b}^t \tilde{A}e = \tilde{b}_2 \tilde{a}_{21} \quad (2.15)$$

The polynomial (2.14) is realizable (in the sense that it corresponds to a 2-stage Runge–Kutta method (2.13)) if the first and last equations in (2.15) can be solved for \tilde{a}_{21} and \tilde{b}_2 in \mathbb{R} . A simple computation gives that a necessary condition is $\beta_{11}^2 - 2\beta_2 \geq 0$. Note that the stability function is independent of α .

With the help of Maple, we have computed the largest r such that (2.14) is a.m. at $(-r, -r)$ and the polynomial is realizable. We have obtained that the optimal perturbation, denoted by \tilde{K}_L , satisfies $\tilde{b}_2 = \tilde{a}_{21} = 0$ and $\tilde{b}_1 = \frac{1}{3}(\sqrt{7} - 2)$, and

$$R_{Lin}^{\text{opt}}(K) = \frac{1}{3}(1 + \sqrt{7}) \approx 1.21525. \quad (2.16)$$

Observe that $R_{Lin}^{\text{opt}}(K) < \tilde{R}_{2,2} = \sqrt{2}$. The stability function (2.1) for the optimal perturbed method is

$$\tilde{\phi}_{(K, \tilde{K}_L)}(z, \tilde{z}) = \frac{1}{9}(4 + \sqrt{7})\left(1 + \frac{z}{r}\right)^2 + \frac{1}{9}(5 - \sqrt{7})\left(1 + \frac{\tilde{z}}{r}\right),$$

where $r = R_{Lin}^{\text{opt}}(K)$.

EXAMPLE 2.4. We consider perturbations of the classical four stage, order four method, of the form

$$\begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ \hline K & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array} \quad \begin{array}{c|cccc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \tilde{a}_{31} & 0 & 0 & 0 \\ \tilde{a}_{41} & \tilde{a}_{42} & 0 & 0 \\ \hline \tilde{K} & \tilde{b}_1 & \tilde{b}_2 & 0 & 0 \end{array} \quad (2.17)$$

We consider these perturbations because, in order to obtain a nonzero SSP coefficient for nonlinear problems, the analysis done in [9] shows that only the entries \tilde{a}_{31} , \tilde{a}_{41} , \tilde{a}_{42} , \tilde{b}_1 and \tilde{b}_2 in \tilde{K} need be nonzero. To study SSP coefficients for the linear case, we have to analyze the perturbed stability function, that in this case is of the form

$$\phi_{(K, \tilde{K})}(z, \tilde{z}) = 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \frac{1}{24}z^4 + \beta_1(z + \tilde{z}) + \beta_{11}z(z + \tilde{z}) + \beta_{21}z^2(z + \tilde{z}) \quad (2.18)$$

where

$$\beta_1 = \tilde{b}_1 + \tilde{b}_2, \quad \beta_{11} = \frac{1}{6}(3\tilde{b}_2 + 2\tilde{a}_{31} + \tilde{a}_{41} + \tilde{a}_{42}), \quad \beta_{21} = \frac{1}{12}(2\tilde{a}_{31} + \tilde{a}_{42}).$$

After some computations with (2.18), we obtain a coefficient $R_{Lin}(K, \tilde{K}) \approx 1.66728$ that is the positive root of the polynomial $15x^4 - 4x^3 - 12x^2 - 24x - 24 = 0$. The coefficients are

$$\beta_1 = \frac{7r^3 - 2r^2 - 6r - 12}{12}, \quad \beta_{11} = \frac{5r^2 - 2r - 6}{12}, \quad \beta_{21} = \frac{r - 1}{6}.$$

where $r = R_{Lin}(K, \tilde{K})$. With these values, the perturbed stability function can be written as

$$\phi_{(K, \tilde{K})}(z, \tilde{z}) = \gamma_{01}\left(1 + \frac{\tilde{z}}{r}\right) + \gamma_{11}\left(1 + \frac{z}{r}\right)\left(1 + \frac{\tilde{z}}{r}\right) + \gamma_{21}\left(1 + \frac{z}{r}\right)^2\left(1 + \frac{\tilde{z}}{r}\right) + \gamma_{40}\left(1 + \frac{z}{r}\right)^4,$$

where

$$\gamma_{01} = \frac{r(2r^3 - r^2 - 6)}{6}, \quad \gamma_{11} = \frac{r^2(r^2 + 2r - 6)}{12}, \quad \gamma_{21} = \frac{r^3(r - 1)}{6}, \quad \gamma_{40} = \frac{r^4}{24}.$$

This perturbed stability function can be realized with the family of perturbations

$$\begin{aligned}\tilde{a}_{31} &= \frac{1}{2} (2r - 2 - \tilde{a}_{42}) , \\ \tilde{a}_{41} &= \frac{1}{2} (5r^2 - 6r - 2 - 6\tilde{b}_2) , \\ \tilde{b}_1 &= \frac{1}{12} (7r^3 - 2r^2 - 6r - 12 - 12\tilde{b}_2) .\end{aligned}$$

3. Perturbed Runge–Kutta methods for nonlinear problems. In this section we seek to answer the question posed in the introduction: for a given method K , what perturbation \tilde{K} gives the largest value of $R(K, \tilde{K})$? We begin by providing some upper bounds.

It is convenient to write scheme (1.7) in canonical Shu–Osher form [5]

$$Y = v_r u_n + \alpha_r \left(Y + \frac{h}{r} F \right) \quad (3.1)$$

where

$$v_r = (I + rK)^{-1} e, \quad \alpha_r = r(I + rK)^{-1} K. \quad (3.2)$$

Observe that matrices K and α_r have the same structure (strictly lower triangular, lower triangular or full).

A Runge–Kutta method (1.7) is said to be absolutely monotonic at r if $(I + rK)^{-1}$ exists and all entries of α_r, v_r are non-negative. For a given method K , the largest r such that the method is absolutely monotonic at r is known as the SSP coefficient, Kraaijevanger coefficient, or radius of absolute monotonicity [18], and it is denoted herein by $R(K)$:

$$R(K) = \sup \{ r \mid r = 0 \text{ or } r > 0, (I + rK)^{-1} \text{ exists, and } \alpha_r, v_r \geq 0 \} . \quad (3.3)$$

As usual, the inequalities above should be understood component–wise.

Next, we consider perturbed Runge–Kutta methods (1.8). To study absolute monotonicity of perturbed Runge–Kutta methods, we write method (1.8) also in a canonical Shu–Osher-like form

$$Y = \gamma_r u_n + \alpha_r^{\text{up}} \left(Y + \frac{h}{r} F \right) + \alpha_r^{\text{down}} \left(Y - \frac{h}{r} \tilde{F} \right), \quad (3.4)$$

where

$$\gamma_r = (I + rK + 2r\tilde{K})^{-1} e, \quad (3.5a)$$

$$\alpha_r^{\text{up}} = r(I + rK + 2r\tilde{K})^{-1} (K + \tilde{K}), \quad (3.5b)$$

$$\alpha_r^{\text{down}} = r(I + rK + 2r\tilde{K})^{-1} \tilde{K}. \quad (3.5c)$$

Observe that method (3.4), with $\gamma_r = (I - \alpha_r^{\text{up}} - \alpha_r^{\text{down}})e$, is a perturbed Runge–Kutta scheme with Butcher coefficients

$$K = \frac{1}{r} (I - \alpha_r^{\text{up}} - \alpha_r^{\text{down}})^{-1} (\alpha_r^{\text{up}} - \alpha_r^{\text{down}}), \quad \tilde{K} = \frac{1}{r} (I - \alpha_r^{\text{up}} - \alpha_r^{\text{down}})^{-1} \alpha_r^{\text{down}}, \quad (3.6)$$

provided that $(I - \alpha_r^{\text{up}} - \alpha_r^{\text{down}})^{-1}$ exists.

The *radius of absolute monotonicity* of a perturbed Runge–Kutta method (K, \tilde{K}) is the largest r such that γ_r , α_r^{up} and α_r^{down} in (3.5) exist and are non-negative [9, Definition 3.1]:

$$R(K, \tilde{K}) = \sup \left\{ r \mid r = 0 \text{ or } r > 0, (I + rK + 2r\tilde{K})^{-1} \text{ exists, and } \gamma_r, \alpha_r^{\text{up}}, \alpha_r^{\text{down}} \geq 0 \right\}. \quad (3.7)$$

3.1. Zero-well-defined perturbations. Regularity of $(I - \alpha_r^{\text{up}} - \alpha_r^{\text{down}})$ is evidently important in our study. Observe that from (3.5) we have

$$(I - \alpha_r^{\text{up}} - \alpha_r^{\text{down}})(I + rK) = (I - 2\alpha_r^{\text{down}}). \quad (3.8)$$

Consequently, if $I + rK$ is regular for some r , then $(I - \alpha_r^{\text{up}} - \alpha_r^{\text{down}})$ is regular if and only if $(I - 2\alpha_r^{\text{down}})$ is regular.

If $I - 2\alpha_r^{\text{down}}$ is singular, then the stage equations do not have a unique solution even for the trivial ODE given by $f = 0$. Hence we say that methods for which $I - 2\alpha_r^{\text{down}}$ is singular are not *zero-well-defined*. See [5, Chap. 3] for the analogous definition in the context of traditional Runge–Kutta methods.

3.2. Optimal perturbations. The *optimal perturbed SSP coefficient* of a Runge–Kutta method K is denoted by

$$R^{\text{opt}}(K) = \sup_{\tilde{K}} R(K, \tilde{K}).$$

For a given method K that is (explicit/diagonally implicit/fully implicit), we consider the supremum over perturbations \tilde{K} that are zero-well-defined and correspond to the same class of methods. A matrix \tilde{K} such that $R(K, \tilde{K}) = R^{\text{opt}}(K)$ is called an optimal perturbation. For some methods, the optimal perturbation is not unique.

The following result shows that every method can be perturbed so as to give a method with strictly positive SSP coefficient.

THEOREM 3.1. *Let K be a Runge-Kutta method that belongs to a specified class of methods (explicit, diagonally implicit, or fully implicit). Then it is always possible to find a perturbation \tilde{K} within the same class such that $R(K, \tilde{K}) > 0$.*

Proof. From [9, Proposition 3.7], we have $R(K, \tilde{K}) > 0$ if and only if the Butcher coefficients satisfy

$$K + \tilde{K} \geq 0, \quad \tilde{K} \geq 0, \quad (3.9)$$

and the following inequalities hold,

$$\text{Inc}((K + 2\tilde{K})(K + \tilde{K})) \leq \text{Inc}(K + \tilde{K}), \quad (3.10a)$$

$$\text{Inc}((K + 2\tilde{K})\tilde{K}) \leq \text{Inc}(\tilde{K}), \quad (3.10b)$$

where $\text{Inc}(F)$ denotes the incidence matrix of matrix F defined as $\text{Inc}(F) = (g_{ij})$ where $g_{ij} = 1$ if $f_{ij} \neq 0$, and $g_{ij} = 0$ if $f_{ij} = 0$.

Consider first the implicit case. By making all entries of \tilde{K} positive we can satisfy (3.10), and by making them large enough we can satisfy (3.9). For the explicit and diagonally implicit cases,

note that if K, \tilde{K} are (strictly) lower-triangular, then the left-hand sides of (3.10) are also. Thus by making all the (strictly) lower-triangular entries of \tilde{K} positive, and by taking them large enough, we can satisfy the above inequalities. \square

Observe that a perturbed Runge–Kutta method (K, \tilde{K}) can be interpreted as an additive Runge–Kutta method $(K + \tilde{K}, \tilde{K})$ for functions (f, \tilde{f}) , and conditions (3.5) are the ones required for the absolute monotonicity of this additive scheme at $(z, \tilde{z}) = (-r, -r)$ (see [10]). From Lemma 2.8 in [10], we obtain that the stability function $\phi_{(K, \tilde{K})}$ defined by (2.1), is absolutely monotonic at $(\xi, \tilde{\xi}) = (-r, -r)$. Consequently,

$$R(K, \tilde{K}) \leq R_{\text{Lin}}(K, \tilde{K}) \leq R_{\text{Lin}}^{\text{opt}}(K). \quad (3.11)$$

Furthermore, from SSP theory and inequality (2.4), we have

$$R(K) \leq R(\phi_K) \leq R_{\text{Lin}}^{\text{opt}}(K), \quad R(K) \leq R^{\text{opt}}(K) \leq R_{\text{Lin}}^{\text{opt}}(K). \quad (3.12)$$

The following example illustrates that $R(\phi_K)$ can be either larger or smaller than $R^{\text{opt}}(K)$.

EXAMPLE 3.2. *We consider the family of second order 2-stage Runge-Kutta methods (2.13) for $\alpha \in \mathbb{R}$. For this family we have*

$$\begin{aligned} v_r \geq 0 & \iff \alpha > 0 \quad \text{and} \quad 0 \leq r \leq \frac{1}{\alpha}, \\ \alpha_r \geq 0 & \iff \alpha \geq \frac{1}{2} \quad \text{and} \quad 0 \leq r \leq \frac{2\alpha - 1}{\alpha}. \end{aligned}$$

Thus

$$R(K) = \begin{cases} 0, & \text{if } \alpha \leq \frac{1}{2}, \\ \frac{2\alpha - 1}{\alpha}, & \text{if } \frac{1}{2} < \alpha \leq 1, \\ \frac{1}{\alpha}, & \text{if } 1 < \alpha. \end{cases} \quad (3.13)$$

In Figure 3.1 we show the threshold factor $R(\phi_K)$ (thin solid blue line) and the SSP coefficient $R(K)$ (thick solid black line). We also show the corresponding optimal coefficients for perturbed methods, namely, the optimal threshold factor $R_{\text{Lin}}^{\text{opt}}(K)$ (thin dashed blue line) and the optimal SSP coefficient $R^{\text{opt}}(K)$ for the perturbed method (thick dashed black line).

We see that for optimal SSP method ($\alpha = 1$) it is not possible to increase the SSP coefficient by means of perturbations. However, for $\alpha = (\sqrt{7} - 1)/2$ it is possible to obtain a perturbation that raises the SSP coefficient to $R^{\text{opt}}(K) = R_{\text{Lin}}^{\text{opt}}(K) = (1 + \sqrt{7}) \approx 1.21525$ (see (2.16)).

We also see that for $2/3 < \alpha < 1$ we obtain $R(\phi_K) < R^{\text{opt}}(K)$, whereas for $0 < \alpha < 2/3$ and for $1 < \alpha$ we have $R^{\text{opt}}(K) < R(\phi_K)$.

Coefficients of the perturbations that give rise to these values are given in Appendix 5.1.

3.3. Upper bounds on the SSP coefficient for perturbed RK methods. In this section, we explore some upper bounds on the SSP coefficient $R^{\text{opt}}(K)$ where K is an s -stage order p method. A straightforward upper bound is obtained from inequality (3.12) and Theorem 2.9:

$$R^{\text{opt}}(K) \leq \sqrt[s]{s(s-1)\dots(s-p+1)}. \quad (3.14)$$

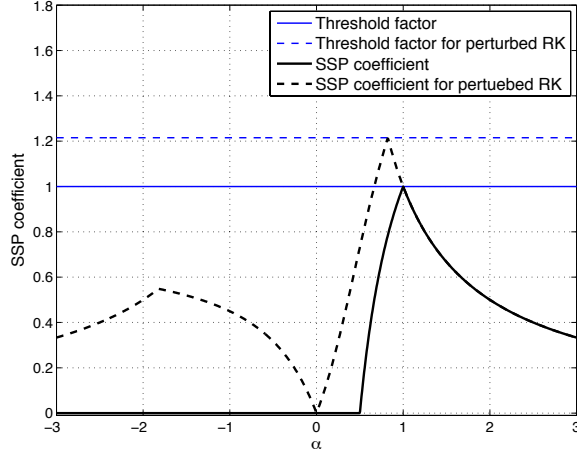


FIG. 3.1. Family of second order 2-stage methods: SSP coefficients for unperturbed methods and optimal SSP coefficients for perturbed methods.

As the next Theorem shows, the largest positive value such that vector v_r in (3.2) is non-negative is also an upper bound for $R^{\text{opt}}(K)$.

THEOREM 3.3. *Consider an explicit Runge-Kutta method K and let r_e be the largest positive value such that vector v_r in (3.2) is non-negative. Then*

$$R^{\text{opt}}(K) \leq r_e. \quad (3.15)$$

Proof. Let $r = R^{\text{opt}}(K)$. Then $\gamma_r = (I - \alpha^{\text{up}} - \alpha^{\text{down}})e \geq 0$, and thus from (3.2) and (3.8) we get

$$(I - 2\alpha_r^{\text{down}})v_r \geq 0. \quad (3.16)$$

As $\alpha_r^{\text{down}} \geq 0$, and since we consider only explicit, zero-well-defined perturbations, $I - 2\alpha_r^{\text{down}}$ is an M matrix. Thus $(I - 2\alpha_r^{\text{down}})^{-1} \geq 0$. If we multiply (3.16) by $(I - 2\alpha_r^{\text{down}})^{-1}$ we obtain that $v_r \geq 0$. \square

From Theorem 3.3 we obtain that

$$R(K) \leq R^{\text{opt}}(K) \leq r_e. \quad (3.17)$$

Consequently, for those methods such that $R(K) = r_e$, the SSP coefficient cannot be increased by perturbation. This is the case for the family of second-order two-stage methods. For $\alpha \geq 1$, $R(K) = r_e = 1/\alpha$ (see Example 3.2).

On the other hand, if $R(K) < r_e$ one can try to find a perturbation to increase the SSP coefficient. This is the case for the classical 4-stage order 4 method for which $R(K) = 0$ and $r_e \approx 1.2956$, the real root of $x^3 - 2x^2 + 4x - 4 = 0$.

Another interesting bound, for explicit methods only, can be obtained in terms of the Butcher coefficients of the Runge-Kutta method K .

THEOREM 3.4. *Consider an explicit Runge-Kutta method (K, \tilde{K}) with perturbed SSP coefficient $R^{\text{opt}}(K) > 0$. Let $K = (a_{ij})$. Then*

$$R^{\text{opt}}(K) \leq \frac{1}{\max_{ij} |a_{ij}|} \quad (3.18)$$

Proof. The proof is similar to that of [22, Lemma 3.2]. Consider an optimal perturbation \tilde{K} and set $r = R^{\text{opt}}(K, \tilde{K}) > 0$; consider too the canonical representation (3.4). Let $\Lambda = \alpha_r^{\text{up}} + \alpha_r^{\text{down}} = (\alpha_{ij})$, $\Gamma = \alpha_r^{\text{up}}/r = (\beta_{ij})$, $\tilde{\Gamma} = \alpha_r^{\text{down}}/r = (\tilde{\beta}_{ij})$; observe that $\Lambda, \Gamma, \tilde{\Gamma} \geq 0$, and that $\Lambda = r(\Gamma + \tilde{\Gamma})$. As $(I - \Lambda)e = \gamma_r \geq 0$ and $\alpha_{ik} \geq 0$, we have $\alpha_{ik} \leq 1$; as $(I - \Lambda)K = \Gamma - \tilde{\Gamma}$, we have

$$a_{ik} = \beta_{ik} - \tilde{\beta}_{ik} + \sum_{j=k+1}^{i-1} \alpha_{ij} a_{jk}. \quad (3.19)$$

As $\alpha_{ik} = r(\beta_{ik} + \tilde{\beta}_{ik})$, then $\beta_{ik} + \tilde{\beta}_{ik} = \alpha_{ik}/r \leq 1/r$. In particular, from (3.19),

$$|a_{21}| = |\beta_{21} - \tilde{\beta}_{21}| \leq \beta_{21} + \tilde{\beta}_{21} \leq \frac{1}{r}.$$

We proceed by induction on row ℓ of K . Assume that $|a_{ij}| \leq 1/r$, for $i = 2, \dots, \ell$, $j = 1, \dots, \ell - 1$, and consider row $\ell + 1$. Then, from (3.19),

$$\begin{aligned} |a_{\ell+1,1}| &= \left| \beta_{\ell+1,1} - \tilde{\beta}_{\ell+1,1} + \sum_{j=2}^{\ell} \alpha_{\ell+1,j} a_{j,1} \right| \leq \beta_{\ell+1,1} + \tilde{\beta}_{\ell+1,1} + \sum_{j=2}^{\ell} \alpha_{\ell+1,j} |a_{j,1}| \\ &\leq \frac{1}{r} \alpha_{\ell+1,1} + \frac{1}{r} \sum_{j=2}^{\ell} \alpha_{\ell+1,j} \leq \frac{1}{r} \sum_{j=1}^{\ell} \alpha_{\ell+1,j} \leq \frac{1}{r}. \end{aligned}$$

A similar argument can be used to show that $|a_{\ell+1,j}| \leq 1/r$, $j = 2, \dots, \ell$. The Theorem follows by induction. \square

Consequently,

$$R(K) \leq R^{\text{opt}}(K) \leq \frac{1}{\max_{ij} |a_{ij}|}.$$

For those methods such that $R(K) = 1/\max_{ij} |a_{ij}|$ it is not possible to increase the SSP coefficient by perturbing the method. This is the case for all known optimal explicit SSP RK methods of orders one through four, with any number of stages [13].

For the restricted class of perturbations considered in [22], similar results were obtained in [22, Theorems 3.1, 3.4, 3.5 and 3.6]. Theorem 3.4 extends those results, showing that no improvement in the radius of absolute monotonicity is possible for many optimal SSP methods, even when more general perturbations are considered.

3.4. Relations among the Butcher and canonical Shu-Osher representations. For perturbations of explicit Runge-Kutta methods (or of other methods with one stage equal to u_n) there exists a certain simple transformation that may yield a larger value of $R(K, \tilde{K})$ – and that never yields a smaller value.

PROPOSITION 3.5. *Let an s -stage explicit perturbed Runge–Kutta method be given with coefficients $\gamma_r, \alpha_r^{\text{up}}, \alpha_r^{\text{down}} \geq 0$, where r is the radius of absolute monotonicity of the method. Consider the perturbed method with coefficients*

$$\hat{\gamma} = (1, 0, \dots, 0)^t, \quad (3.20a)$$

$$\hat{\alpha}_{i,1}^{\text{up}} = \alpha_{i,1}^{\text{up}} + (\gamma_r)_i/2 \quad 2 \leq i \leq s \quad (3.20b)$$

$$\hat{\alpha}_{i,1}^{\text{down}} = \alpha_{i,1}^{\text{down}} + (\gamma_r)_i/2 \quad 2 \leq i \leq s. \quad (3.20c)$$

Then the perturbed method with coefficients $(\gamma_r, \alpha_r^{\text{up}}, \alpha_r^{\text{down}})$ and the modified perturbed method with coefficients $(\hat{\gamma}, \hat{\alpha}^{\text{up}}, \hat{\alpha}^{\text{down}})$ correspond to the same RK method K . The modified perturbed method has radius of absolute monotonicity at least equal to r .

Proof. It is easily seen that the modified method is equivalent to the original one when $f = \tilde{f}$, so they correspond to the same unperturbed method. Meanwhile, the transformation never leads to negative coefficients, so the modified method is a.m. at r . \square

REMARK 2. *Proposition 3.5 is also valid for Runge–Kutta methods whose first row is equal to zero.*

In the Butcher form (1.8) it is obvious which perturbed methods (K, \tilde{K}) correspond to a given method K . In the canonical Shu–Osher form it is less obvious. The following lemma characterizes which methods of the form (3.4) are perturbations of a given method (3.1).

LEMMA 3.6. *If method (3.4) is a perturbation of method (3.1), then their coefficients are related as follows:*

$$(I - 2\alpha_r^{\text{down}})\alpha_r = (\alpha_r^{\text{up}} - \alpha_r^{\text{down}}) \quad (3.21a)$$

$$(I - 2\alpha_r^{\text{down}})v_r = \gamma_r. \quad (3.21b)$$

Furthermore, if (3.21) holds and the perturbation is zero-well-defined, then (3.4) is a perturbation of (3.1).

Proof. To prove the first part, take $\tilde{f} = f$ in (3.4) to obtain:

$$Y = \gamma_r u_n + (\alpha_r^{\text{up}} + \alpha_r^{\text{down}})Y + (\alpha_r^{\text{up}} - \alpha_r^{\text{down}})\frac{h}{r}F.$$

Subtract $2\alpha_r^{\text{down}}Y$ from both sides to get

$$(I - 2\alpha_r^{\text{down}})Y = \gamma_r u_n + (\alpha_r^{\text{up}} - \alpha_r^{\text{down}})\left(Y + \frac{h}{r}F\right). \quad (3.22)$$

Substituting (3.1) in the above gives

$$(I - 2\alpha_r^{\text{down}})v_r u_n + (I - 2\alpha_r^{\text{down}})\alpha_r \left(Y + \frac{h}{r}F\right) = \gamma_r u_n + (\alpha_r^{\text{up}} - \alpha_r^{\text{down}})\left(Y + \frac{h}{r}F\right),$$

Equating coefficients yields (3.21).

To prove the second part, assume $I - 2\alpha_r^{\text{down}}$ is invertible and write (3.21) as

$$\alpha_r = (I - 2\alpha_r^{\text{down}})^{-1}(\alpha_r^{\text{up}} - \alpha_r^{\text{down}}) \quad (3.23a)$$

$$v_r = (I - 2\alpha_r^{\text{down}})^{-1}\gamma_r. \quad (3.23b)$$

Substitute (3.23) in (3.1), multiply on the left by $(I - 2\alpha^{\text{down}})^{-1}$, and follow the steps above in reverse. \square

Lemma 3.6 does *not* imply that the perturbation (3.4) is unique for a given r ; see Proposition 3.5.

REMARK 3. *The necessity of the zero-well-defined condition in the second part of Proposition 3.6 can be seen from the following example. We take the implicit trapezoidal Runge-Kutta method*

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1/2 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array}.$$

The canonical form (3.1) is then

$$\alpha_r = \begin{pmatrix} 0 & 0 & 0 \\ \frac{r}{r+2} & \frac{r}{r+2} & 0 \\ \frac{r}{r+2} & \frac{r}{r+2} & 0 \end{pmatrix}, \quad v_r = \begin{pmatrix} 1 \\ \frac{2-r}{r+2} \\ \frac{2-r}{r+2} \end{pmatrix}.$$

Then (3.21) is satisfied – for any r – by

$$\alpha^{\text{up}} = \alpha^{\text{down}} = \begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 1/2 & 0 \end{pmatrix}, \quad \gamma = \begin{pmatrix} 1/3 \\ 0 \\ 0 \end{pmatrix}.$$

However, this method – which involves a perturbation that is not zero-well-defined – is not a perturbation of the original method. \square

3.5. Computing optimal perturbations. In this section we present two algorithms to symbolically or numerically find the optimal perturbed SSP coefficient and a corresponding perturbation of a given RK method. The first algorithm is proven to approximate the optimal value to any accuracy, contingent on the computational solution of linear program subproblems. It is only valid for explicit perturbations. The second algorithm is analytical and exact, and valid for both explicit and implicit methods, but it is not proven to give the optimal value. The results of the two algorithms coincide (to high precision) for all explicit methods on which we have tested them.

3.5.1. Provably optimal algorithm for explicit perturbations. In the foregoing, we have shown that finding an optimal perturbation consists of determining the largest r such that there exists a splitting satisfying (3.21) with positive coefficients. Note that the range of values for which a method (K, \tilde{K}) is absolutely monotonic is always the interval $[0, R(K, \tilde{K})]$. Therefore, one way to find the largest r is to devise a method for testing for a given r whether there exists a perturbation \tilde{K} such that $R(K, \tilde{K}) \geq r$. For given method (1.7) and value of r , the system of equations (3.21) together with the inequalities $\alpha_r^{\text{up}}, \alpha_r^{\text{down}} \geq 0$ constitutes a linear programming (LP) feasibility problem. The following theorem is an immediate consequence of Lemma 3.6.

THEOREM 3.7. *Let an s -stage RK method K and a positive number r be given. There exists a perturbation \tilde{K} with $R(K, \tilde{K}) \geq r$ if and only if there exists an $(s+1) \times (s+1)$ matrix α_r^{down} such that $(I - 2\alpha_r^{\text{down}})$ is regular and the following componentwise inequalities hold:*

$$(I - 2\alpha_r^{\text{down}})\alpha_r + \alpha_r^{\text{down}} \geq 0 \tag{3.24a}$$

$$(I - 2\alpha_r^{\text{down}})v_r \geq 0 \tag{3.24b}$$

$$\alpha_r^{\text{down}} \geq 0. \tag{3.24c}$$

The linear program (3.24) can be solved by standard LP solvers. By embedding this solution in a one-dimensional root-finding algorithm, optimal perturbations can be found. An algorithm based on bisection follows.

Algorithm 1 Optimal explicit perturbation

Input: K

```

 $r_{\max} := 1 / \max |a_{ij}|, r_{\min} := 0.$ 
while  $r_{\max} - r_{\min} > \epsilon$  do
   $r = \frac{r_{\max} + r_{\min}}{2}.$ 
  Compute the coefficient matrices  $\alpha_r, v_r$  using (3.2).
  Solve the LP given by (3.24).
  if it is feasible then
     $r_{\min} := r$ 
  else
     $r_{\max} := r$ 
  end if
end while
return  $r_{\min}$ 

```

Assuming the solution of the LP is correct, the algorithm provably finds an optimal explicit perturbation. However, for implicit perturbations the LP solver may converge to a solution (like the method in Remark 3 above) for which $I - 2\alpha_r^{\text{down}}$ is singular.

3.5.2. Iterated splitting algorithm. We next investigate how to choose $\alpha_r^{\text{up}}, \alpha_r^{\text{down}}$ directly so as to find a perturbation with radius of a.m. at least r . The following result suggests an approach.

LEMMA 3.8. *Given an explicit Runge–Kutta method (3.1), let $\alpha_r^{\text{up}} \geq 0, \alpha_r^{\text{down}} \geq 0$ denote coefficients of a zero-well-defined perturbation of (3.1). Then there exist matrices $\alpha^+ \geq 0, \alpha^- \geq 0$ such that*

$$\alpha_r^{\text{up}} = (I + 2\alpha^-)^{-1} \alpha^+ \quad (3.25a)$$

$$\alpha_r^{\text{down}} = (I + 2\alpha^-)^{-1} \alpha^-, \quad (3.25b)$$

and $\alpha_r = \alpha^+ - \alpha^-$.

Proof. Since the perturbation is zero-well-defined, we can define

$$\alpha^+ = (I - 2\alpha_r^{\text{down}})^{-1} \alpha_r^{\text{up}} \quad (3.26a)$$

$$\alpha^- = (I - 2\alpha_r^{\text{down}})^{-1} \alpha_r^{\text{down}}. \quad (3.26b)$$

Then, by (3.21a), $\alpha_r = \alpha^+ - \alpha^-$. Furthermore, since $I - 2\alpha_r^{\text{down}}$ is an M -matrix, we have $\alpha^+ \geq 0$ and $\alpha^- \geq 0$. Solving (3.26) for $\alpha_r^{\text{up}}, \alpha_r^{\text{down}}$ gives (3.25). \square

In the next algorithm we use the following notation:

$$((x)^+)_{ij} = \begin{cases} x_{ij} & \text{if } x_{ij} \geq 0 \\ 0 & \text{if } x_{ij} < 0. \end{cases} \quad ((x)^-)_{ij} = \begin{cases} 0 & \text{if } x_{ij} \geq 0 \\ -x_{ij} & \text{if } x_{ij} < 0, \end{cases}$$

and thus $x = (x)^+ - (x)^-$ is a sign splitting of matrix x , with $(x)^+ \geq 0, (x)^- \geq 0$.

Given a perturbed Runge-Kutta method (3.4) with $\gamma_r = e_1$, where $e_1 = (1, 0, \dots, 0)^t$, and α^{up} or α^{down} containing negative values, we construct

$$\tilde{\gamma}_r = (I + 2(\alpha_r^{\text{up}})^- + 2(\alpha_r^{\text{down}})^-)^{-1} e_1, \quad (3.27a)$$

$$\tilde{\alpha}_r^{\text{up}} = (I + 2(\alpha_r^{\text{up}})^- + 2(\alpha_r^{\text{down}})^-)^{-1} ((\alpha_r^{\text{up}})^+ + (\alpha_r^{\text{down}})^-), \quad (3.27b)$$

$$\tilde{\alpha}_r^{\text{down}} = (I + 2(\alpha_r^{\text{up}})^- + 2(\alpha_r^{\text{down}})^-)^{-1} ((\alpha_r^{\text{up}})^- + (\alpha_r^{\text{down}})^+), \quad (3.27c)$$

where $\alpha^{\text{up}} = (\alpha^{\text{up}})^+ - (\alpha^{\text{up}})^-$, $\alpha^{\text{down}} = (\alpha^{\text{down}})^+ - (\alpha^{\text{down}})^-$, provided that $I + 2(\alpha_r^{\text{up}})^- + 2(\alpha_r^{\text{down}})^-$ exists. Using Lemma 3.6, it is straightforward to prove that, if method α_r^{up} , α_r^{down} is a perturbation of (3.1), then (3.27) is also perturbation of (3.1). Next, for explicit methods, we perform transformation (3.20). In this way, (3.27) followed by transformation (3.20) give a perturbation of the form (3.4) with $\gamma_r = e_1$, that we denote by $\hat{\alpha}_r^{\text{up}}$, $\hat{\alpha}_r^{\text{down}}$. If $\hat{\alpha}_r^{\text{up}} \geq 0$, $\hat{\alpha}_r^{\text{down}} \geq 0$, then r is an SSP coefficient; otherwise, we can repeat the above process.

The following lemma studies the sign of $(\hat{\alpha}_r^{\text{up}})_{ij}$, $(\hat{\alpha}_r^{\text{down}})_{ij}$ when $(\alpha_r^{\text{up}})_{ij} < 0$ or $(\alpha_r^{\text{down}})_{ij} < 0$. For the sake of clarity, we drop the index r .

LEMMA 3.9. *We consider a perturbed explicit Runge-Kutta method with coefficients $\gamma = e_1$, α^{up} , α^{down} , and the perturbation $\hat{\alpha}^{\text{up}}$, $\hat{\alpha}^{\text{down}}$ obtained by computing (3.27) followed by transformation (3.20). Assume that $j_0 \geq 2$ is the first row with negative terms in α^{up} or α^{down} . Let m_0 be the largest index $m_0 \geq 1$ such that $\alpha_{j_0, m_0}^{\text{up}} < 0$ or $\alpha_{j_0, m_0}^{\text{down}} < 0$. Then*

1. *For first to $(j_0 - 1)$ -th row, we have: $\hat{\alpha}_{i,j}^{\text{up}} = \alpha_{i,j}^{\text{up}}$ and $\hat{\alpha}_{i,j}^{\text{down}} = 0$ for $1 \leq i \leq j_0 - 1$, $1 \leq j \leq j_0 - 2$.*
2. *For the j_0 -th row, we have:*
 - (a) *If $m_0 = 1$, then $\hat{\alpha}_{j_0,1}^{\text{up}} < 0$ or $\hat{\alpha}_{j_0,1}^{\text{down}} < 0$.*
 - (b) *If $m_0 \geq 2$, then, $\hat{\alpha}_{j_0, m_0}^{\text{up}} \geq 0$ and $\hat{\alpha}_{j_0, m_0}^{\text{down}} \geq 0$.*
 - (c) *For $1 \leq m_0 \leq j_0 - 2$, we have $\hat{\alpha}_{j_0, \ell}^{\text{up}} \geq 0$ and $\hat{\alpha}_{j_0, \ell}^{\text{down}} \geq 0$ for $\ell = m_0 + 1, \dots, j_0 - 1$.*

Proof. If j_0 is the first row with negative terms in α^{up} or α^{down} , straightforward computations give that $\hat{\alpha}_{i,j}^{\text{up}} = \alpha_{i,j}^{\text{up}}$ and $\hat{\alpha}_{i,j}^{\text{down}} = 0$ for $1 \leq i \leq j_0 - 1$, $1 \leq j \leq j_0 - 2$, and

$$\hat{\alpha}_{j_0,1}^{\text{up}} = \alpha_{j_0,1}^{\text{up}} - 2 \sum_{i=2}^{j_0-1} ((\alpha_{j_0,i}^{\text{up}})^- + (\alpha_{j_0,i}^{\text{down}})^-) \alpha_{i,1}^{\text{up}}, \quad (3.28a)$$

$$\hat{\alpha}_{j_0,\ell}^{\text{up}} = (\alpha_{j_0,\ell}^{\text{up}})^+ + (\alpha_{j_0,\ell}^{\text{down}})^- - 2 \sum_{i=\ell+1}^{j_0-1} ((\alpha_{j_0,i}^{\text{up}})^- + (\alpha_{j_0,i}^{\text{down}})^-) \alpha_{i,\ell}^{\text{up}}, \quad \ell = 2, \dots, j_0 - 1. \quad (3.28b)$$

and

$$\hat{\alpha}_{j_0,1}^{\text{down}} = \alpha_{j_0,1}^{\text{down}} - 2 \sum_{i=2}^{j_0-1} ((\alpha_{j_0,i}^{\text{up}})^- + (\alpha_{j_0,i}^{\text{down}})^-) \alpha_{i,1}^{\text{down}}, \quad (3.29a)$$

$$\hat{\alpha}_{j_0,\ell}^{\text{down}} = (\alpha_{j_0,\ell}^{\text{up}})^- + (\alpha_{j_0,\ell}^{\text{down}})^+ - 2 \sum_{i=\ell+1}^{j_0-1} ((\alpha_{j_0,i}^{\text{up}})^- + (\alpha_{j_0,i}^{\text{down}})^-) \alpha_{i,\ell}^{\text{down}}, \quad \ell = 2, \dots, j_0 - 1. \quad (3.29b)$$

Let m_0 be the largest index $m_0 \geq 1$ such that $\alpha_{j_0, m_0}^{\text{up}} < 0$ or $\alpha_{j_0, m_0}^{\text{down}} < 0$. In this case, $\alpha_{j_0, i}^{\text{up}} \geq 0$,

$\alpha_{j_0,i}^{\text{down}} \geq 0$ for $i = m_0 + 1, \dots, j_0 - 1$, and thus

$$(\alpha_{j_0,i}^{\text{up}})^+ = \alpha_{j_0,i}^{\text{up}}, \quad (\alpha_{j_0,i}^{\text{down}})^+ = \alpha_{j_0,i}^{\text{down}}, \quad (\alpha_{j_0,i}^{\text{up}})^- = (\alpha_{j_0,i}^{\text{down}})^- = 0, \quad i = m_0 + 1, \dots, j_0 - 1.$$

If $m_0 = 1$, from (3.28a) and (3.29a) we get $\hat{\alpha}_{j_0,1}^{\text{up}} = \alpha_{j_0,1}^{\text{up}}$ and $\hat{\alpha}_{j_0,1}^{\text{down}} = \alpha_{j_0,1}^{\text{down}}$, and thus $\hat{\alpha}_{j_0,1}^{\text{up}} < 0$ or $\hat{\alpha}_{j_0,1}^{\text{down}} < 0$. If $m_0 \geq 2$, from (3.28b) and (3.29b) we get

$$\hat{\alpha}_{j_0,m_0}^{\text{up}} = (\alpha_{j_0,m_0}^{\text{up}})^+ + (\alpha_{j_0,m_0}^{\text{down}})^- \geq 0, \quad \hat{\alpha}_{j_0,m_0}^{\text{down}} = (\alpha_{j_0,m_0}^{\text{up}})^- + (\alpha_{j_0,m_0}^{\text{down}})^+ \geq 0.$$

Finally, for $1 \leq m_0 \leq j_0 - 2$, from (3.28b) and (3.29b) we get that, for $\ell = m_0 + 1, \dots, j_0 - 1$, we have

$$\hat{\alpha}_{j_0,\ell}^{\text{up}} = (\alpha_{j_0,\ell}^{\text{up}})^+ \geq 0, \quad \hat{\alpha}_{j_0,\ell}^{\text{down}} = (\alpha_{j_0,\ell}^{\text{down}})^+ \geq 0.$$

□

Consequently, if matrices α_r^{up} and α_r^{down} contain negative elements in the second or later columns, an iterated construction of perturbations $\hat{\alpha}_r^{\text{up}}$, $\hat{\alpha}_r^{\text{down}}$ removes these negative values obtaining a perturbation with non-negative elements from second column on. However, if in a row j_0 we have:

$$\alpha_{j_0,1}^{\text{up}} < 0 \quad \text{and} \quad \alpha_{j_0,\ell}^{\text{up}} \geq 0 \quad \ell = 2, \dots, j_0 - 1, \quad (3.30)$$

or

$$\alpha_{j_0,1}^{\text{down}} < 0 \quad \text{and} \quad \alpha_{j_0,\ell}^{\text{down}} \geq 0 \quad \ell = 2, \dots, j_0 - 1, \quad (3.31)$$

the new perturbation $\hat{\alpha}_r^{\text{up}}$, $\hat{\alpha}_r^{\text{down}}$ will also contain negative elements in the first column.

We now give Algorithm 2 to determine whether there exists a perturbation with a.m. radius r for a given method.

Algorithm 2 Existence of a perturbation with radius r

Input: r, K

Compute the coefficient matrices α_r, v_r using (3.2).

Set $\alpha^{\text{up}} = \alpha_r$ and $\alpha^{\text{down}} = 0$.

while α^{up} or α^{down} has any negative entries **do**

 If K has a zero row, perform the transformation (3.20).

 If $\alpha^{\text{up}}, \alpha^{\text{down}} \geq 0$, stop. This is a feasible perturbation.

 If condition (3.30) or (3.31) hold, stop. A feasible perturbation cannot be found.

 Set $\alpha^- = (\alpha^{\text{up}})^- + (\alpha^{\text{down}})^+$ and $\alpha^+ = (\alpha^{\text{up}})^+ + (\alpha^{\text{down}})^-$

 Compute a new splitting:

$$\begin{aligned} \alpha^{\text{up}} &= (I + 2((\alpha^{\text{up}})^- + (\alpha^{\text{down}})^-))^{-1} \alpha^+ \\ \alpha^{\text{down}} &= (I + 2((\alpha^{\text{up}})^- + (\alpha^{\text{down}})^-))^{-1} \alpha^- \end{aligned}$$

end while

Order	Stages	Method	$R(K)$	$R^{\text{opt}}(K)$	Bound (3.18)	Bound (3.14)	Property C
1	1	Forward Euler	1	1	1	1	True
2	2	Midpoint	0	0.732	1	1.414	True
	2	Minimal trunc. error	0.5	1	1.333	1.414	True
	2	SSP22 [23]	1	1	1	1.414	True
	2	SSP22* [6]	0.784	1.215	1.215	1.414	True
3	3	Heun33 [8]	0	0.776	1.333	1.817	False
	3	SSP33 [23]	1	1	1	1.817	True
4	4	RK44 (Kutta)	0	0.685	1	2.213	False
	5	Merson [19]	0	0.242	0.5	3.309	False
	10	SSP104 [13]	6	6	6	8.425	False
5	6	Fehlberg [4]	0	0.057	0.125	3.727	False
	7	Dormand-Prince [3]	0	0.040	0.086	4.789	False
	8	Bogacki [1]	0	0.313	0.859	5.827	False
	7	SSP75 [22]	0	1.396	1.792	4.789	False
	8	SSP85 [22]	0	1.875	1.919	5.827	True
	9	SSP95 [22]	0	2.738	3.198	6.853	False
6	9	Calvo [2]	0	0.021	0.059	6.265	False
8	13	Prince-Dormand [20]	0	0.013	0.059	9.212	False

TABLE 3.1

Properties of some RK methods and their optimal perturbations. The optimal perturbed radius of absolute monotonicity was computed by both the linear programming algorithm and the iterated splitting algorithm; in every case they gave identical results (up to roundoff errors). Decimal values have been truncated to the number of digits shown.

REMARK 4. *This approach seems to lead to optimal splittings for all the explicit methods on which we have tested it. However, for all implicit methods we have tested, it fails to increase the radius of absolute monotonicity at all. Even for explicit methods, we have no proof that it's optimal because one could use $(\alpha_r)^+ + \delta, (\alpha_r)^- + \delta$, in place of $(\alpha_r)^+, (\alpha_r)^-$, where δ is any non-negative matrix.* \square

3.6. Examples. In this section we compute optimal perturbations of some existing methods, using the algorithms described in the last section.

We have computed optimal perturbations for several known explicit methods using the two algorithms described above. In all cases, the two algorithms gave the same values. It thus seems possible that Algorithm 2 also gives truly optimal results in general, but we do not have a proof. Properties of the methods studied are given in Table 3.1. Several interesting facts are evident:

- For all optimal SSP methods (up to order four), perturbation cannot yield a larger coefficient. This is evident already from the bound (3.18). For all other methods, some improvement is achieved.
- Consistent with Theorem 3.4, for every method considered, it is possible to achieve $R^{\text{opt}} > 0$ by some perturbation.
- The simple bound (3.18) predicts the optimal coefficient to within a factor of three in every

case.

- The methods SSP75, SSP85, and SSP95 are optimal methods found in [22], with property C. By considering methods without property C, we obtain slightly larger coefficients for perturbations of SSP75 and SSP95. On the other hand, relaxing the column assumption gives no benefit in the case of the SSP85 method.
- The values found have been truncated to three decimal places but are known to greater precision. For the 4-stage, order-four method of Kutta, the three-digit value of $R^{\text{opt}}(K)$ given in the table matches the value found by Shu and Osher. However, the exact (irrational) value is slightly larger and is given in the appendix.

4. Conclusions. In this work we have studied SSP coefficients for perturbations of a given explicit Runge-Kutta method. We have considered both the linear and the nonlinear case, and have obtained useful bounds on the threshold factor and on the radius of absolute monotonicity for perturbed Runge-Kutta methods. We have also provided an algorithm for computing optimal perturbations of explicit Runge-Kutta methods, and given optimal perturbations for many methods from the literature.

This work seems to provide a complete picture for the case of most interest: explicit methods applied to nonlinear problems. Nevertheless, some other interesting issues remain unsolved. These include:

- A method to compute optimal perturbations for linear problems.
- An algorithm for obtaining optimal splittings of implicit methods.

These may be a starting point for future work.

5. Appendix. In this section we give additional details on SSP coefficients and optimal perturbations of second order 2-stage Runge-Kutta methods and the classical 4-stage fourth order Runge-Kutta method.

5.1. Second order 2-stage methods. We consider the family of 2-stage second order methods (2.13). In example 2.3 we studied perturbations that increase the SSP coefficient for the linear case. For nonlinear problems, in example 3.2, figure 3.1 shows the values of $R^{\text{opt}}(K)$ for $\alpha \in [-3, 3]$.

In this section, for each α , we give the expressions for $R^{\text{opt}}(K)$ and we show optimal perturbations \tilde{K}_{NL} such that $R(K, \tilde{K}_{NL}) = R^{\text{opt}}(K)$. It is important to point out the convenience of choosing $\tilde{K}_{NL} = \tilde{K}_L$, where \tilde{K}_L denotes the optimal perturbation for the linear case. In this case, we have not only $R(K, \tilde{K}_L) = R^{\text{opt}}(K)$ but also $R_{\text{Lin}}(K, \tilde{K}_{NL}) = R_{\text{Lin}}^{\text{opt}}(K)$. The computations required to obtain the results in this section have been done with the symbolic computation program *Mathematica*.

If we denote by $r = R^{\text{opt}}(K)$, we have that

$$r = \begin{cases} \frac{1}{|\alpha|}, & \text{if } \alpha \in \left(-\infty, -\frac{1}{2}(1 + \sqrt{7})\right] \cup \left[\frac{1}{2}(-1 + \sqrt{7}), \infty\right), \\ \frac{-1 + \alpha + \sqrt{3\alpha^2 - 2\alpha + 1}}{|\alpha|}, & \text{if } \alpha \in \left(-\frac{1}{2}(1 + \sqrt{7}), 0\right) \cup \left(0, \frac{1}{2}(-1 + \sqrt{7})\right). \end{cases} \quad (5.1)$$

Next we give optimal perturbations \tilde{K}_{NL} .

For $\alpha < 0$, we obtain that it is not possible to obtain a perturbation of the form (2.13) with $\tilde{b}_2 = 0$ and $\tilde{a}_{21} = 0$. Consequently, $\tilde{K}_{NL} \neq \tilde{K}_L$ and we always have that $R_{\text{Lin}}(K, \tilde{K}_{NL}) < R_{\text{Lin}}^{\text{opt}}(K)$.

Optimal perturbations of the form (2.13) for different values of $\alpha < 0$ must satisfy the following conditions.

- For $-\frac{1}{2}(1 + \sqrt{7}) \leq \alpha < 0$, the coefficients \tilde{a}_{21} , \tilde{b}_1 and \tilde{b}_2 in \tilde{K}_{NL} must satisfy

$$-\alpha \leq \tilde{a}_{21} \leq \frac{1-r\alpha}{2r}, \quad \tilde{b}_1 = -\frac{r\tilde{a}_{21}}{2\alpha}, \quad \tilde{b}_2 = -\frac{1}{2\alpha},$$

where $r = R^{\text{opt}}(K)$.

- For $\alpha \leq -\frac{1}{2}(1 + \sqrt{7})$, we should have

$$\tilde{a}_{21} = -\alpha, \quad -\frac{1}{2\alpha} \leq \tilde{b}_1 \leq \frac{-2\alpha^2 - 2\alpha + 1}{4\alpha}, \quad -\frac{1}{2\alpha} \leq \tilde{b}_2 \leq \frac{2\alpha\tilde{b}_1 - 1}{4\alpha}.$$

For $\alpha > 0$ we can find optimal perturbations with $\tilde{b}_2 = 0$ and $\tilde{a}_{21} = 0$. Coefficient \tilde{b}_1 must satisfy the following conditions.

- For $0 < \alpha \leq (-1 + \sqrt{7})/2$, we have that

$$\tilde{b}_1 = \frac{\sqrt{3\alpha^2 - 2\alpha + 1} - \alpha}{2\alpha}. \quad (5.2)$$

Thus there is a unique \tilde{K}_{NL} of the form (2.13). In this case, we have $R(K) < R(K, \tilde{K}_{NL}) = R^{\text{opt}}(K)$.

- For $(-1 + \sqrt{7})/2 < \alpha < 1$, we also get $R(K) < R^{\text{opt}}(K)$, but in this case the optimal perturbation \tilde{K}_{NL} is not unique. All the perturbations with \tilde{b}_1 satisfying

$$\frac{1-\alpha}{\alpha} \leq \tilde{b}_1 \leq \frac{2\alpha^2 - 2\alpha + 1}{4\alpha},$$

are optimal. In particular, we can take $\tilde{K}_{NL} = \tilde{K}_L$. With this choice, $R(K, \tilde{K}_L) = R^{\text{opt}}(K) = 1/\alpha$ and $R_{\text{Lin}}(K, \tilde{K}_L) = R_{\text{Lin}}^{\text{opt}}(K) \approx 1.22$. Furthermore, $\alpha = (-1 + \sqrt{7})/2$ provides the largest SSP coefficient within the family of 2-stage second order method (see figure 3.1).

- For $1 \leq \alpha$, we have $R(K) = R^{\text{opt}}(K) = 1/\alpha$ and the optimal perturbation \tilde{K}_{NL} is not unique. All the values

$$0 \leq \tilde{b}_1 \leq \frac{2\alpha^2 - 2\alpha + 1}{4\alpha}$$

give optimal perturbations. We can take $\tilde{K}_{NL} = 0$, but in this case $R_{\text{Lin}}(K, 0) < R_{\text{Lin}}^{\text{opt}}(K)$.

A better choice is $\tilde{K}_{NL} = \tilde{K}_L$. Observe that, for $\alpha = 1$, we get the optimal SSP coefficient $R(K) = 1$ that cannot be increased by perturbations.

Next, we consider some concrete values of α to show the the expressions of the perturbations. For each value, we give the Butcher tableau of the perturbation and matrices α^{up} and α^{down} in (3.4).

- For $\alpha = 1/2$ we get method RK2a in [12] with $R(K) = 0$. With perturbation

$$\tilde{K} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \tilde{b}_1 & 0 & 0 \end{pmatrix}, \quad \alpha^{\text{up}} = \begin{pmatrix} 0 & 0 & 0 \\ \tilde{b}_1 & 0 & 0 \\ 0 & 2\tilde{b}_1 & 0 \end{pmatrix}, \quad \alpha^{\text{down}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 - 2\tilde{b}_1 & 0 & 0 \end{pmatrix}, \quad \gamma = \begin{pmatrix} 1 \\ 1 - \tilde{b}_1 \\ 0 \end{pmatrix},$$

where $\tilde{b}_1 = \frac{1}{2}(\sqrt{3} - 1)$, we get $R(K, \tilde{K}) = R_{\text{Lin}}(K, \tilde{K}) = \sqrt{3} - 1$.

- For $\alpha = 2/3$, we have a nontrivial SSP coefficient $R^{\text{opt}}(K) = 1/2$, but we can increase this value to $R(K, \tilde{K}_1) = R^{\text{opt}}(K) = 1$ with perturbation

$$\tilde{K}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1/4 & 0 & 0 \end{pmatrix}, \alpha^{\text{up}} = \begin{pmatrix} 0 & 0 & 0 \\ 2/3 & 0 & 0 \\ 0 & 3/4 & 0 \end{pmatrix}, \alpha^{\text{down}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1/4 & 0 & 0 \end{pmatrix}, \gamma = \begin{pmatrix} 1 \\ 1/3 \\ 0 \end{pmatrix}.$$

For this perturbation, $R(\phi_K) = R_{\text{Lin}}(K, \tilde{K}_1) = 1$. We can take $\gamma = (1, 0, 0)^t$ by modifying the first column of α^{up} and α^{down} according to (3.20),

$$\tilde{K}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 1/6 & 0 & 0 \\ 3/8 & 0 & 0 \end{pmatrix}, \alpha^{\text{up}} = \begin{pmatrix} 0 & 0 & 0 \\ 5/6 & 0 & 0 \\ 0 & 3/4 & 0 \end{pmatrix}, \alpha^{\text{down}} = \begin{pmatrix} 0 & 0 & 0 \\ 1/6 & 0 & 0 \\ 1/4 & 0 & 0 \end{pmatrix}, \gamma = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

- As it has been pointed out above, the largest value in the α -family of 2-stage second order schemes is $R^{\text{opt}}(K) = (1 + \sqrt{7})/3$ and it is obtained for $\alpha = (\sqrt{7} - 1)/2$. The perturbation is of the form (2.13) with $\tilde{b}_1 = (\sqrt{7} - 2)/2$, and

$$\alpha^{\text{up}} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{9}(4 + \sqrt{7}) & 0 \end{pmatrix}, \alpha^{\text{down}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{1}{9}(5 - \sqrt{7}) & 0 & 0 \end{pmatrix}, \gamma_r = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

This is the perturbation obtained in [6, Table V] by numerical search in the class of perturbations considered in [6].

5.2. Classical fourth order 4-stage method. For nonlinear problems, applying the analysis above, we find that the optimal perturbation of the classical method has SSP coefficient given by the real root of $x^3 + 2x^2 + 4x - 4 = 0$, which is approximately $R^{\text{opt}}(K) \approx 0.685016$. The corresponding perturbation is *not unique*. For instance, we can take $\gamma_r = (1, 0, 0, 0, 0)$, and all entries of α_r^{down} equal to zero except

$$(\alpha_r^{\text{down}})_{31} = \frac{r^2}{4} \quad (\alpha_r^{\text{down}})_{42} = \frac{r^2}{2}, \quad (5.3)$$

where $r = R^{\text{opt}}(K)$. However, there exist other optimal perturbations with additionally $(\alpha_r^{\text{down}})_{42} = \epsilon$ where $0 \leq \epsilon \leq 0.782$.

We remark that nearly-optimal perturbations for this method are given in [23, p. 448] and [11]. Interestingly, these different perturbed methods have different values of $R_{\text{Lin}}(K, \tilde{K})$.

REFERENCES

- [1] P. Bogacki and Lawrence F. Shampine. An efficient Runge-Kutta (4, 5) pair. *Comput. Math. with Appl.*, 32(6):15–28, 1996.
- [2] M. Calvo, J. I. Montijano, and L. Rández. A new embedded pair of runge-kutta formulas of orders 5 and 6. *Comput. Math. Appl.*, 20(1):15–24, 1990.
- [3] J. R. Dormand and P. J. Prince. A family of embedded Runge-Kutta formulae. *J. Comput. Appl. Math.*, 6(1):19–26, 1980.
- [4] E. Fehlberg. Klassische runge-kutta-formeln fünfter und siebenter ordnung mit schrittweiten-kontrolle. *Computing*, 4(2):93–106, 1969.

- [5] S. Gottlieb, D. I. Ketcheson, and C. W. Shu. *Strong Stability Preserving Runge-Kutta and Multistep Time Discretizations*. World Scientific Publishing Company, 2011.
- [6] S. Gottlieb and S. J. Ruuth. Optimal strong-stability-preserving time-stepping schemes with fast downwind spatial discretizations. *J. Sci. Comput.*, 27:289–303, 2006.
- [7] Y. Hadjimichael and D. I. Ketcheson. Strong stability preserving additive linear multistep methods. In preparation.
- [8] K. Heun. Neue methoden zur approximativen integration der differentialgleichungen einer unabhängigen veränderlichen. *Z. Math. Phys*, 45:23–38, 1900.
- [9] I. Higueras. Representations of Runge-Kutta methods and strong stability preserving methods. *SIAM J. Numer. Anal.*, 43:924–948, 2005.
- [10] I. Higueras. Strong Stability for Additive Runge-Kutta Methods. *SIAM J. Numer. Anal.*, 44(4):1735–1758, 2006.
- [11] I. Higueras. Positivity properties for the classical fourth order Runge-Kutta methods. *Monografías de la Real Academia de Ciencias de Zaragoza*, 33:125–139, 2010.
- [12] W. Hundsdorfer, B. Koren, M. van Loon, and J. C. Verwer. A positive finite-difference advection scheme. *J. Comput. Phys.*, 117(1):35–46, 1995.
- [13] D. I. Ketcheson. Highly Efficient Strong Stability Preserving Runge-Kutta Methods with Low-Storage Implementations. *SIAM J. Sci. Comput.*, 30:2113–2136, 2008.
- [14] D. I. Ketcheson. Computation of optimal monotonicity preserving general linear methods. *Math. Comp.*, 78:1497–1513, 2009.
- [15] D. I. Ketcheson. *High Order Strong Stability Preserving Time Integrators and Numerical Wave Propagation Methods for Hyperbolic PDEs*. Ph. d. thesis, University of Washington, 2009.
- [16] D. I. Ketcheson. Step Sizes for Strong Stability Preservation with Downwind-biased Operators. *SIAM J. Numer. Anal.*, 49(4):1649–1660, 2011.
- [17] D. I. Ketcheson. Nodepy software version 0.6.1, 2015. <http://github.com/ketch/nodepy>.
- [18] J. F. B. M. Kraaijevanger. Contractivity of Runge-Kutta Methods. *BIT*, 31:482–528, 1991.
- [19] R. H. Merson. An operational method for the study of integration processes. In *Proc. Symp. Data Processing*, pages 1–25, 1957.
- [20] P. J. Prince and J. R. Dormand. High order embedded Runge-Kutta formulae. *J. Comput. Appl. Math.*, 7(1):67–75, March 1981.
- [21] S. J. Ruuth. Global optimization of explicit strong-stability-preserving Runge-Kutta Methods. *Math. Comp.*, 75:183–207, 2006.
- [22] S. J. Ruuth and R. J. Spiteri. High-order strong-stability-preserving Runge-Kutta methods with downwind-biased spatial discretizations. *SIAM J. Numer. Anal.*, 42:974–996, 2004.
- [23] C. W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *J. Comput. Phys.*, 77(2):439–471, August 1988.